FloVis: Leveraging Visualization to Protect Sensitive Network Infrastructure

J. Glanfield, D. Paterson, C. Smith, T. Taylor, S. Brooks, C. Gates[ξ], J. McHugh[Ψ]

Dalhousie University
6050 University Ave.
Halifax, Nova Scotia B3H 1W5
Canada

CA Labs[ξ]
One CA Plaza
Islandia, New York, 11749
USA

University of North Carolina[Ψ]
CB 3175, Sitterson Hall
Chapel Hill, NC, 27599
USA

glanfield@cs.dal.ca / paterson@cs.dal.ca / christop@dal.ca / teryl@cs.dal.ca

sbrooks@cs.dal.ca / carrie.gates@ca.com / mchugh@cs.unc.edu

ABSTRACT TITLE

The analysis and evaluation of the network traffic passing over NATO's communication networks is an important yet daunting task for its operators, analysts and system security officers due to the volume and complexity of the data. The following report presents ongoing work in the development of an extensible suite of network traffic visualization tools, called FloVis, that aims to address these issues. FloVis is a visualization framework designed to incorporate different network visualizations seamlessly, as plugins, under one application. The current iteration of the toolset is described below which uses visualization as a mechanism for capturing different aspects of network traffic. Each tool allows the analyst to understand/interpret/model the data in a different but related way. By combining them under the FloVis framework, the analyst has integrated access to each view to help focus his investigation onto areas of interest. As a demonstration of FloVis' effectiveness and its potential, examples have been provided below which illustrate how it can be used for understanding network traffic and detecting security events.

1.0 INTRODUCING THE FLOVIS FRAMEWORK

Computer networks have become critical to NATO operations. Much of NATO's computer traffic runs over civilian networks, and NATO computers are accessible to a wide variety of malicious activities. The scale of the network traffic involved makes monitoring and analysis difficult, and the rapid deployment of computer systems to new areas places additional stresses on operators and analysts. We have developed an extensible suite of visualization tools, FloVis, to aid system administrators and system security officers in understanding the traffic that passes over their networks. The suite is useful for both defensive purposes as well as for evaluating and understanding the effects of offensive information operations. This paper describes FloVis and provides examples of its capabilities. FloVis is a visualization framework that was built with the aim of providing the necessary machinery to allow security analysts to leverage the benefits of data visualization while attempting to detect malicious network behavior [5]. This is accomplished not only by providing new and interesting visualizations, but by allowing these visualizations to synergize their unique perspectives to provide further insight into network data. FloVis was developed to promote:

- Extensibility: The integration of additional visualizations is seamless.
- Inter-visualization communication: Visualizations may communicate without prior knowledge of each other's existence.

In its current state, FloVis consists of a supporting framework and the following plug-ins:

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOV 2010	**N/A**	**-**

4. TITLE AND SUBTITLE
FloVis: Leveraging Visualization to Protect Sensitive Network Infrastructure

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dalhousie University 6050 University Ave Halifax, Nova Scotia B3H 1W5 Canada

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)
11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADA564697. Information Assurance and Cyber Defence (Assurance de l'information et cyberdefense). RTO-MP-IST-091

14. ABSTRACT

The analysis and evaluation of the network traffic passing over NATO's communication networks is an important yet daunting task for its operators, analysts and system security officers due to the volume and complexity of the data. The following report presents ongoing work in the development of an extensible suite of network traffic visualization tools, called FloVis, that aims to address these issues. FloVis is a visualization framework designed to incorporate different network visualizations seamlessly, as plugins, under one application. The current iteration of the toolset is described below which uses visualization as a mechanism for capturing different aspects of network traffic. Each tool allows the analyst to understand/interpret/model the data in a different but related way. By combining them under the FloVis framework, the analyst has integrated access to each view to help focus his investigation onto areas of interest. As a demonstration of FloVis' effectiveness and its potential, examples have been provided below which illustrate how it can be used for understanding network traffic and detecting security events.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**SAR**	**10**	

1.1 Overflow

This visualization provides a high-level overview of network usage by organizations or enterprises that may be hierarchical in nature (see Figures 1 and 2). It focuses on the administrative relationships rather than being network centric as are the other FloVis components. It is purposefully high-level in order to provide motivation for more-detailed analysis of network entities (e.g., hosts or subnets) with more detailed visualizations [1]. In using Overflow, the analyst defines the organizational structure and incorporates the address ranges assigned to each component. The visualization displays inter-organizational traffic patterns and volumes. Unexpected or unusual communication patterns can lead to further investigation of the specific network entities involved.

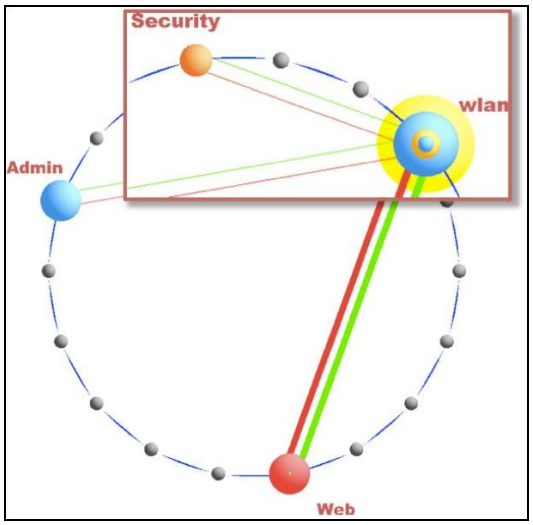

Figure 1: An organizational overview of a network

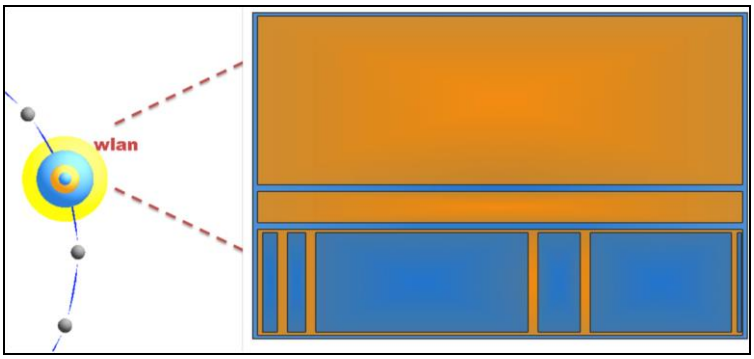

Figure 2: A secondary visualization - an organization

1.2 FlowBundle

Displays communications between network entities and reduces occlusion by using hierarchical edge-bundling [2]. The display arranges 512 entities around a circle that is divided into two sections by the border across which communication was observed (see Figure 3). Typically, the entities are subnet or host addresses with 256 points allocated to each side of the border, but other relationships including host port usage can be displayed. The limitation to 512 points allows unambiguous identification of individual connections; however, a sliding window allows any consecutive 8 bits of the entity ID (e.g., IP address) to

be selected. For example, if the inside network is a single /24, it is possible to view connections between outside /8s and inside hosts. By sliding the outside window, connections from /16s within a given /8 or the /24s within a given /16, etc., can be displayed. Connection line transparency is an approximate indicator of traffic volume. Given that the OverFlow plug-in has identified questionable inter-organizational traffic, FlowBundle could be used to identify the subnets or hosts involved.

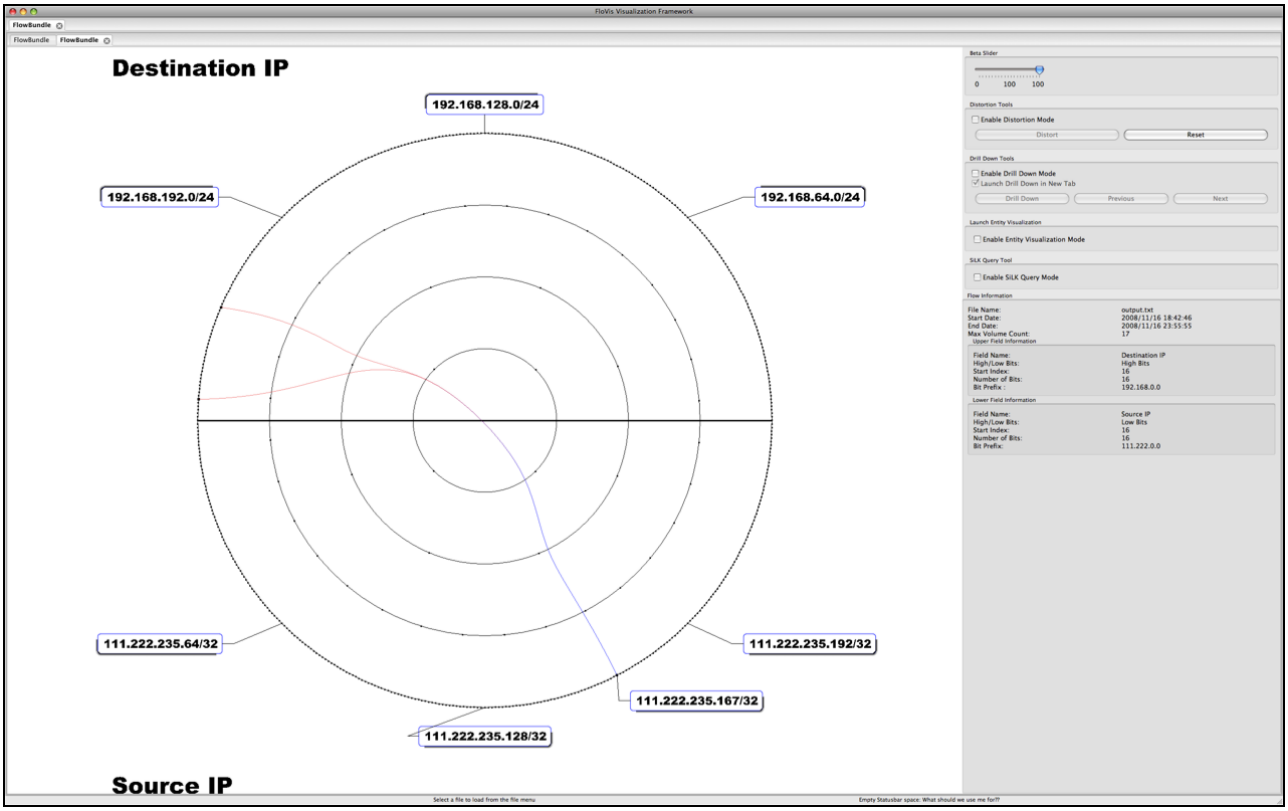

Figure 3: Our host-of-interest communicating with two subnets

1.3 Activity Viewer

This visualization shows categorical entity activity as a function of time, using distinct colors to describe a limited number of categories [5]. The choice of categories is arbitrary. Any small set of behaviors that can be derived from the available data is suitable. One example uses client/server behavior, another shows hosts' responses to scans. The categories of individual entities are plotted against time in a simple two-dimensional grid, with the entities listed along the vertical axis and time along the horizontal axis (see Figure 4). If a given entity exhibits one or more of the categorical activities during a given hour, the corresponding square is given the color of the activity that causes the most concern. In an operational setting, the categories might correspond to the roles assigned to individual hosts. Hosts behaving in manners consistent with their assigned role would be given colors that identify the role and indicate normal activity. Hosts that appear to deviate from the role would be given colors that indicate the nature and extent of the deviation. This could be compressed into the common three category "stoplight chart" with green indicating normal, yellow questionable, and red clearly negative. Since some role shifting and deviation from expectation is often observed, the time series of colors allows quick identification of hosts that are deviating from past behavior. The entities need not be hosts. Subnets or organizational units could be used, as well.

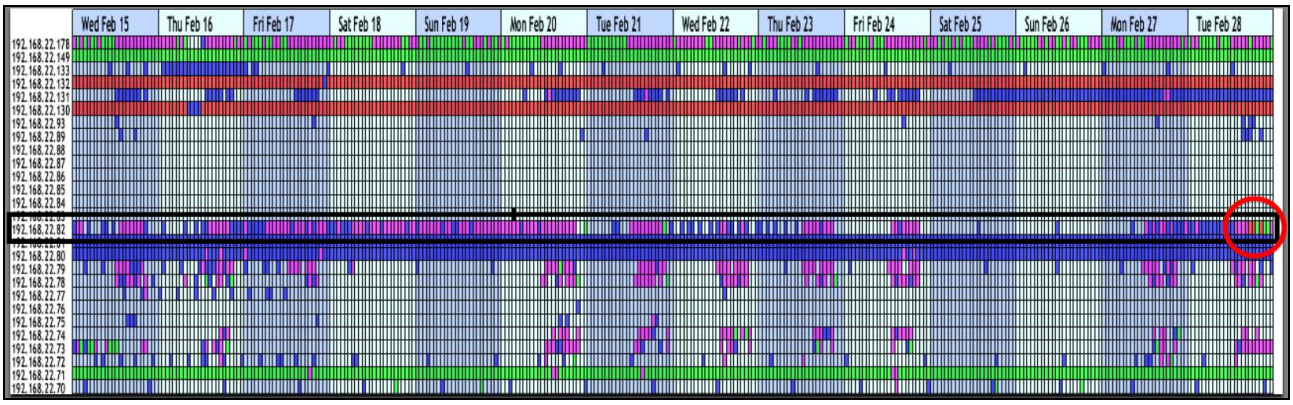

Figure 4: Discovering a change in the pattern of behavior

1.4 NetBytes Viewer

This plug-in allows detailed analysis of host behaviors over time. It displays an impulse plot in three dimensions that describes port or protocol volumes over time [4]. To avoid the occlusion problems that often accompany static 3D plots, the NetBytes Viewer plot can be rotated and rescaled by the user. 2D finder lines allow precise identification of specific impulses in the time/volume and port, protocol/volume planes (see Figures 5 and 8). The viewer is particularly useful in examining the behavior of a compromised machine since the behavior of the machine prior to and after a compromise can easily be compared. In addition, unexpected behavior changes associated with a compromise can be detected. These might include bot behaviors or other malicious activity.

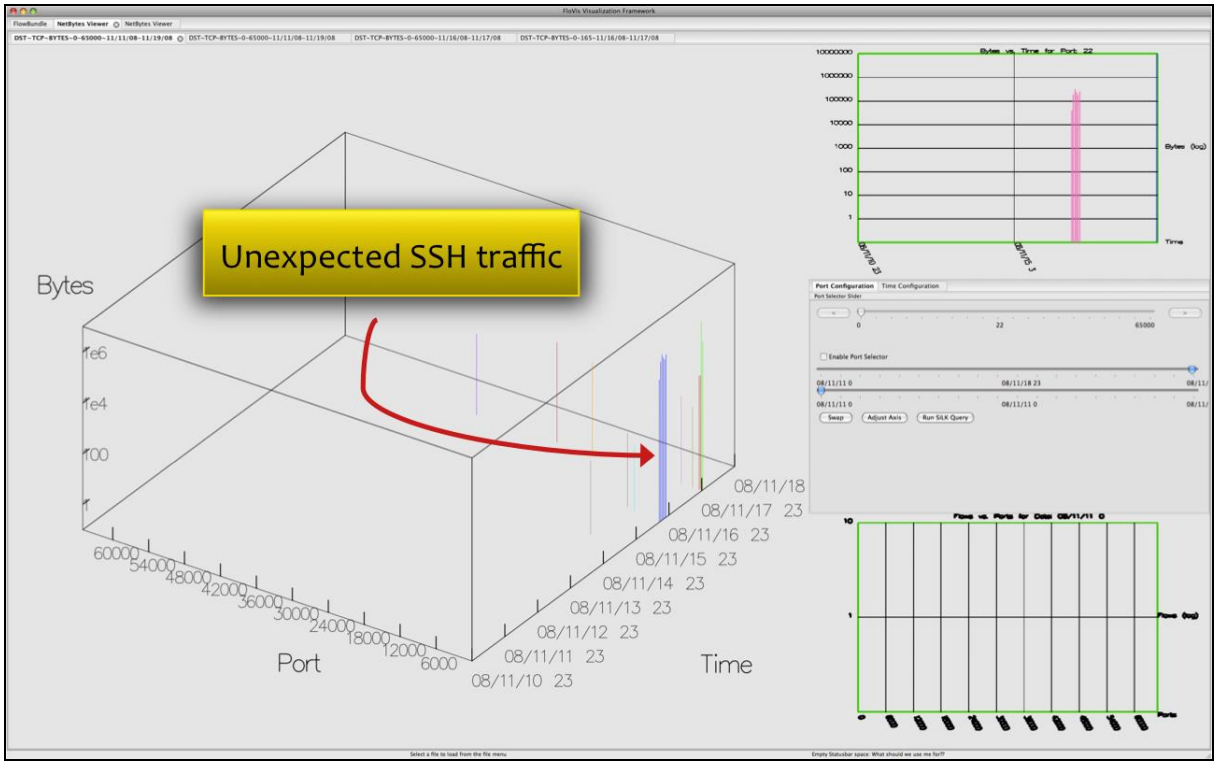

Figure 5: Port 22 traffic occurring over a few hours. The top right of the display shows hundreds of thousands of bytes over a short time period

1.5 Caluster

Caluster is a calendaring visualization (see Figures 9 and 10) (inspired by van Wijk and van Selow, [6]) that displays multiple time series of a quantitative network property, e.g., daily series of port 80 volumes measured hourly. We know that traffic patterns exhibit diurnal patterns, and the basic display allows us to superimpose a large numbers of days of data on a single display, linking each day's line to an accompanying calendar. Clustering algorithms can then be applied to the lines to group them by a variety of similarity measures. When the lines are grouped, each group is assigned a color and the calendar entries for each group are colored accordingly. Once the obvious groupings are accounted for, e.g., weekdays vs. weekend days, smaller groups and singular cases should arise. Among these we hope to find subtle indications of abnormal activities. Since most network activity exhibits fairly strong daily patterns, we suspect that it will be necessary to remove daily patterns before clustering on other time scales is effective.

2.0 CASE STUDIES

In our system, the plug-ins are designed to interact through a number of data properties, enabling both drill down to examine smaller groupings of entities in greater detail as well as pivoting that allows data values of interest in one view to be used as the targets of investigation in others. This interaction is illustrated in the first two case studies. The third case study illustrates the use of the data series clustering tool. While the series in question are time series, the model is more general and can be extended to arbitrary series.

The framework is intended to be extensible so as to accommodate new visualizations easily. Our final case study illustrates this with a new plug-in that has been added to allow visualization of some behavioral properties of flow data.

2.1 Case 1: A subverted system

This example shows how three of the plug-ins offer unique perspectives of the same data. We drill further into the data via each plug-in and discover undesirable traffic.

Using OverFlow to provide a visual breakdown of important subnets, we found unexpected traffic between two organizations (“Security” and “wlan”, see Figure 1) belonging to a conference network.

OverFlow displays a tree-map [3] in order to show volume quantities within levels of a hierarchy. Based on this secondary view of an organization (see Figure 2), it was decided to further investigate the host represented by the large orange block in the tree-map since that block corresponds to the largest portion of traffic. Further drill down with the FlowBundle plug-in reveals that a single host communicating with two subnets, which belong to the “Security” organization (see Figure 3), is responsible for all of the suspicious traffic. The network administrators knew beforehand that there was to be no communication between “Security” and the public network (“wlan”) organizations. Hence, the nature of this traffic is of interest.

The specific nature of the traffic can be determined by using the NetBytes plug-in, as it provides specific port and volume information. With NetBytes, we were able to discover that communication occurred over port 22 and that the volume was large enough to suggest that significant data transfer had occurred (see Figure 5).

In spite of the fact that we drill into the data across multiple visualizations, context is retained through FloVis' multi-tabbed and multi-window displays, which allow the multiple views to be open simultaneously.

2.2 Case 2: Finding anomalous activities

In this example we show how the use of the Activity Viewer allows us to find anomalous behavior by alerting our attention to a change in a host's pattern of behavior. In Figure 4, we see that a host exhibits suspicious server and client activity on the same port. This behavior is of concern because it is unusual for the given host based on the previously observed behavior.

Since we are interested in understanding the specific nature of the traffic occurring on our host of interest, we drill down by using the FlowBundle plug-in. Figures 6 and 7 show us that the host is scanning multiple networks and that it is scanning across a large range of ports, respectively.

The NetBytes plug-in provides yet another perspective by displaying port patterns over time. Thus, we can determine precisely when the scanning activity commenced (see Figure 8).

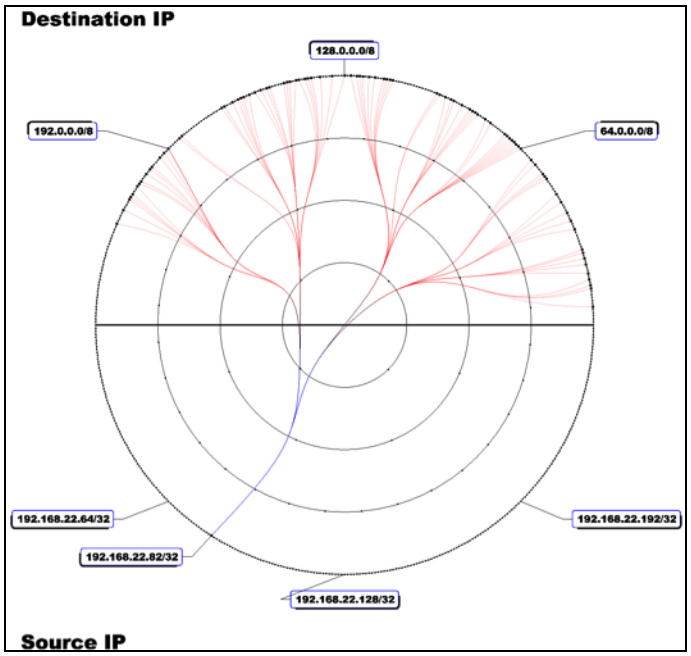

Figure 6: Scanning across networks

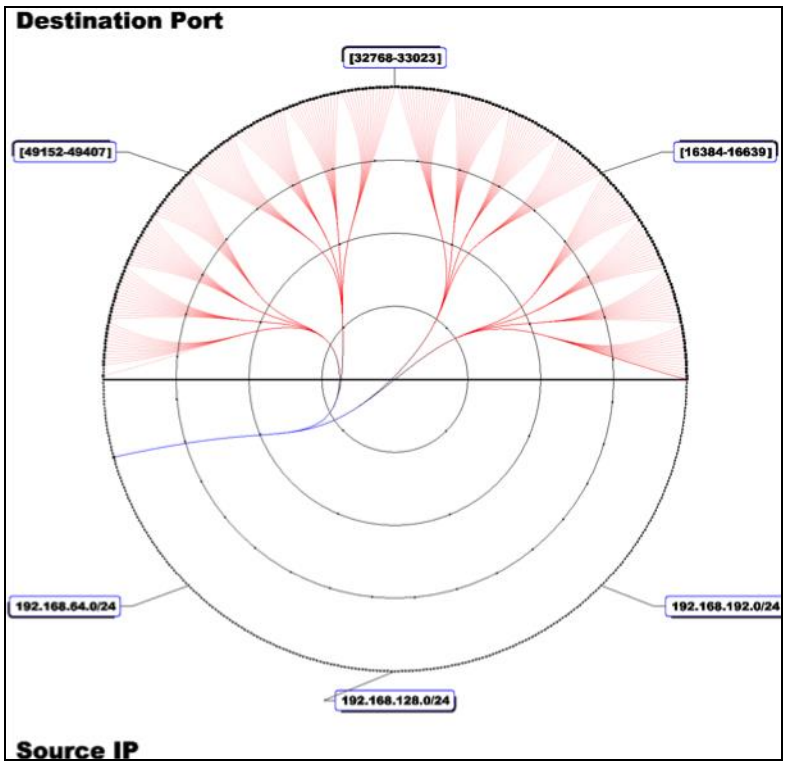

Figure 7 Scanning across ports

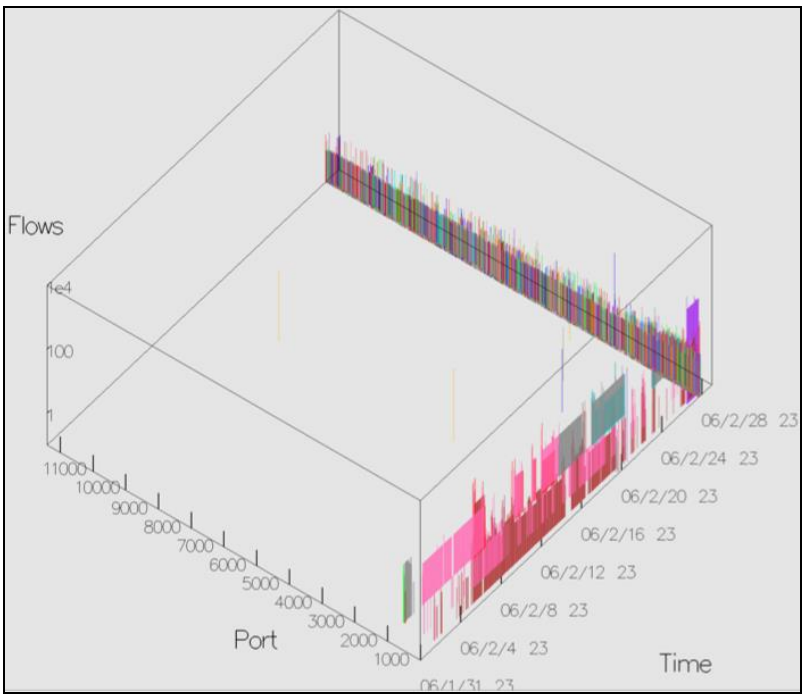

Figure 8 Port-traffic patterns over time

2.3 Case 3: Calendar-based time series clustering

When security analysts investigate large networks for intrusive behaviour, they examine huge datasets (GB+) for signs (or patterns) of anomalous activity. As a result, data reduction is an important function. They must reduce these large datasets into something that is manageable for analysis. There are various ways to reduce data. We can filter it, count it, profile it, and even cluster it. Caluster is a new addition to the FloVis framework that is designed to cluster time series data sets using a visual paradigm first introduced by Jarke J. van Wijk and Edward R. van Selow [6] to cluster the number of employees present in an office building over the day. This technique was adapted to the area of network data analysis in order to find interesting network traffic patterns over time.

The idea behind Caluster is to merge similar daily patterns of some specific network traffic attribute (which could be volumes, connections, etc.) such that similar day patterns are clustered and distinguishable from patterns that are not similar. Caluster does this through two distinct mechanisms. First, it processes daily univariate time series network data through a clustering algorithm in an attempt to find similar day patterns. Secondly, it visualizes those daily patterns in such a way as to draw the users attention to those clusters and also to any anomalous patterns that do not fit in the clusters formed. An example screen shot of Caluster is shown in Figure 9. In this situation, we are clustering the number of unique internal hosts on a small private network that make external HTTP connections on an hourly basis between February 2006 and January 2007. In other words, we are clustering the hourly web surfing patterns of users on a small private network. The data in this case is filtered NetFlow records counted based on destination traffic to port 80. We clustered this data using hierarchical agglomerative clustering (HAC). In HAC, each daily set of data points are treated as their own cluster. During each iteration of the algorithm, the two daily patterns that are most similar (smallest calculated distance function) are merged into a new cluster. The algorithm continues to merge clusters until there is only one cluster left.

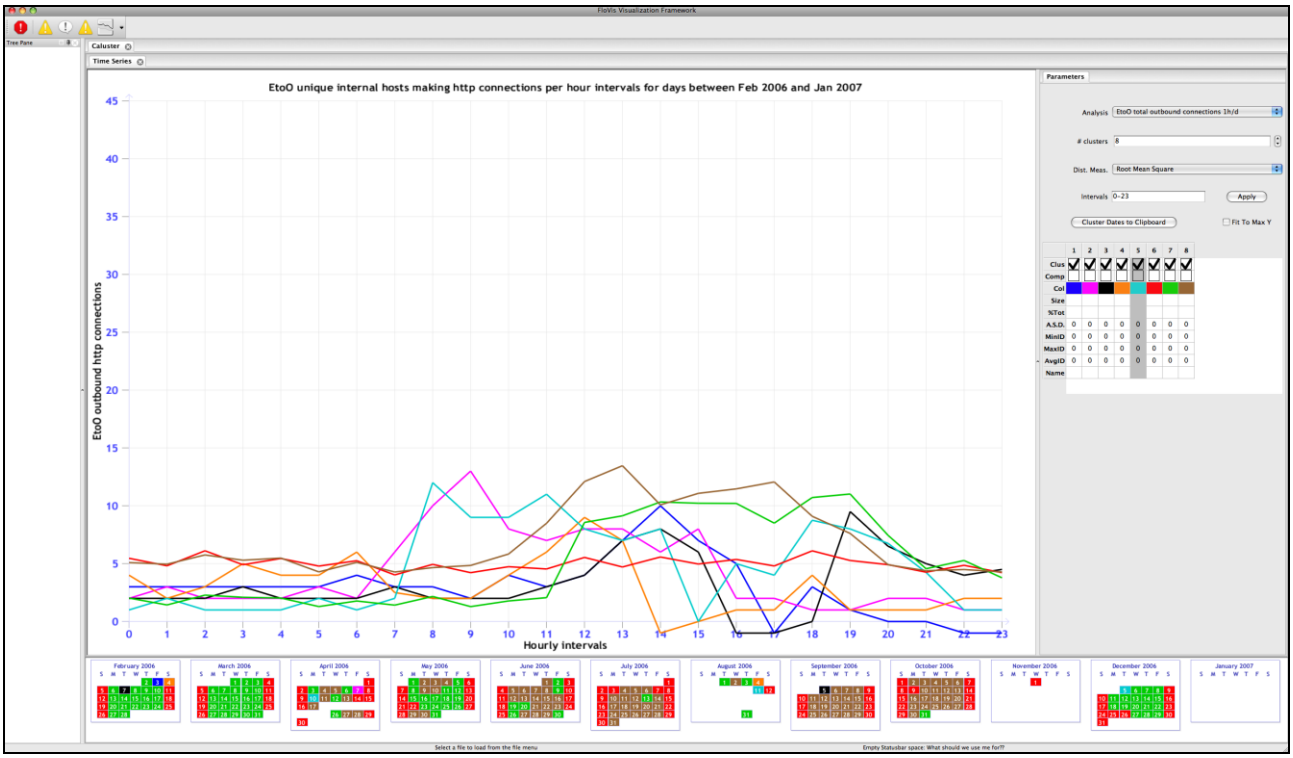

Figure 9: Caluster: Time Series clustering visualization

As shown in Figure 9, there is a calendaring component along the bottom of the interface. Each day in the calendar, represents a day in the dataset. Colour is used to represent the cluster for which a day pattern is merged into. The plot portion of the visualization shows the daily traffic pattern for each cluster with the x-axis representing hourly connection intervals and the y-axis representing the number of unique HTTP connections. As can be seen from the figure, weekends and holidays are typically represented by the red cluster, which shows a relatively constant number of web connections to the outside. Weekdays tend to cluster to the brown or green clusters, which have peak volume times in the afternoon.

Data that does not cluster well is shown in other clusters. For instance, April 7th shows a pink cluster in which web usage spikes early in the day and declines sharply as people go home early. Calendar spots with no data, indicate missing data for the day. Cluster lines that drop down below the x-axis (see the black cluster for hours 16 and 17) indicate missing data points for those hours. We deal with missing data points, by calculating the distance measure between clusters for the existing points and then spreading the distance measure over the full 24-hour period.

Users can arbitrarily choose the number of clusters that they want by using the arrowed "clusters" field on the parameters panel. This allows them to interactively see how clusters form and find a match for the appropriate number of clusters. The application also allows the user to show the data points that compose the cluster. These are shown as stippled lines drawn in the same colour as the cluster line.

In Figure 10, we see the number of unique outbound TCP connections over the time period February 2006 to January 2007. This data does not show the same diurnal patterns as were shown in Figure 9, however, it does show some interesting spikes which deviate from the typical patterns of network traffic. Interestingly, the spikes seem to occur on the 19th and 20th of their respective months. Looking further into the spikes, we see that they are internal hosts that scan external hosts for a single hour (not during work time hours) and then stop. This would definitely raise the suspicions of a security analyst.

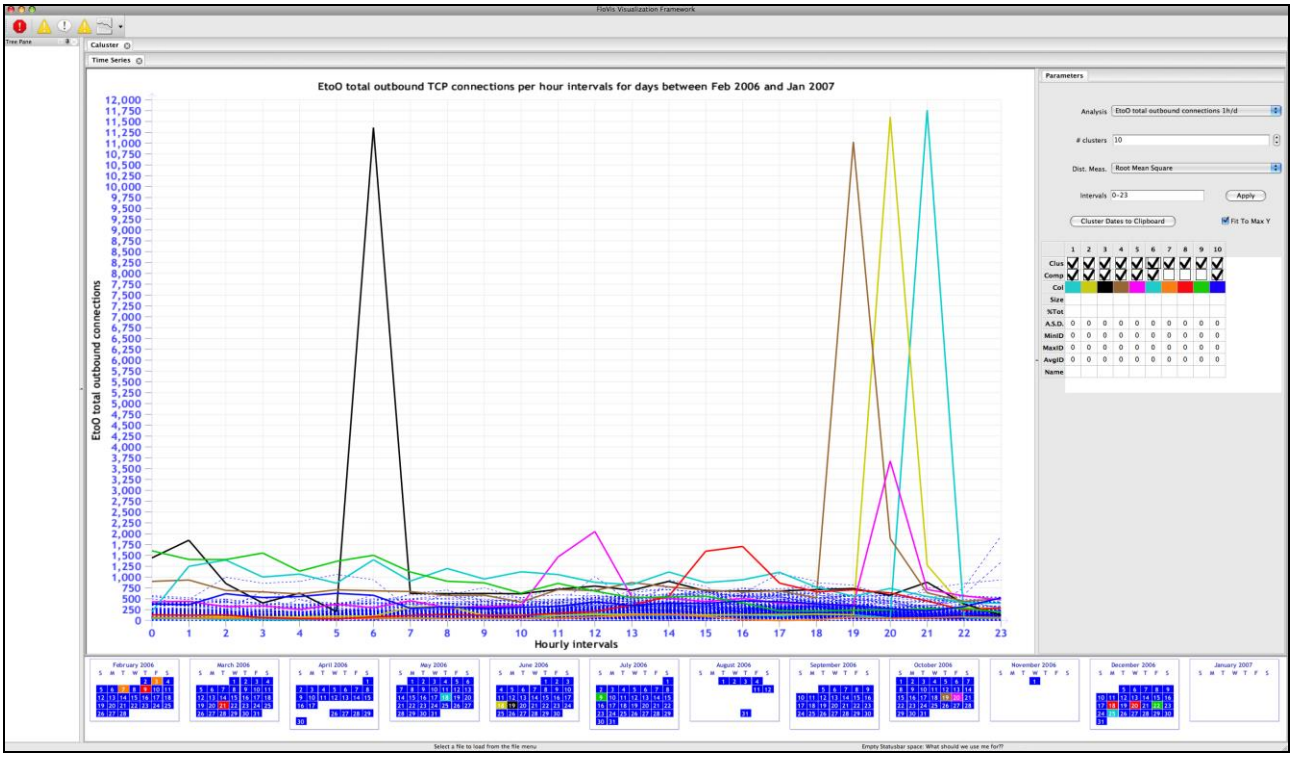

Figure 10: Number of unique internal to external TCP connections

2.4 Case 4: Extending FloVis for Behavioral classification

One of the benefits of using FloVis as a visualization framework is its ease of extensibility. For example, we have recently incorporated some work done at Sonalysts, Inc. Kiayias et al. presented a framework that considers the aggregated behaviors of individual hosts by considering network patterns across multiple hyperplanes [7]. A more detailed exposition of this work is being presented later at this meeting. As it turns out, various views of the feature space of host-behaviors tend to provide insights that are useful to a network analyst. However, the research team at Sonalysts had not as yet leveraged any tools that allow an interactive exploration of the views. We felt it would be helpful to use FloVis to provide such capabilities.

Within the FloVis framework, accomplishing such a feat was relatively easy. Since each visualization is a plug-in to the framework, we simply developed a new plug-in that visualized some of the data used in the research done by the Sonalysts team. Figure 11 shows a visual representation of part of the feature space described above. As can be seen, plotting attributes of flow data can result in interesting clusters of behavior. In this case, the picture describes the number of flows (coming into our monitored network) versus the number of bytes (coming into our monitored network). These axes can be modified to view other flow attributes, or operations on these attributes.

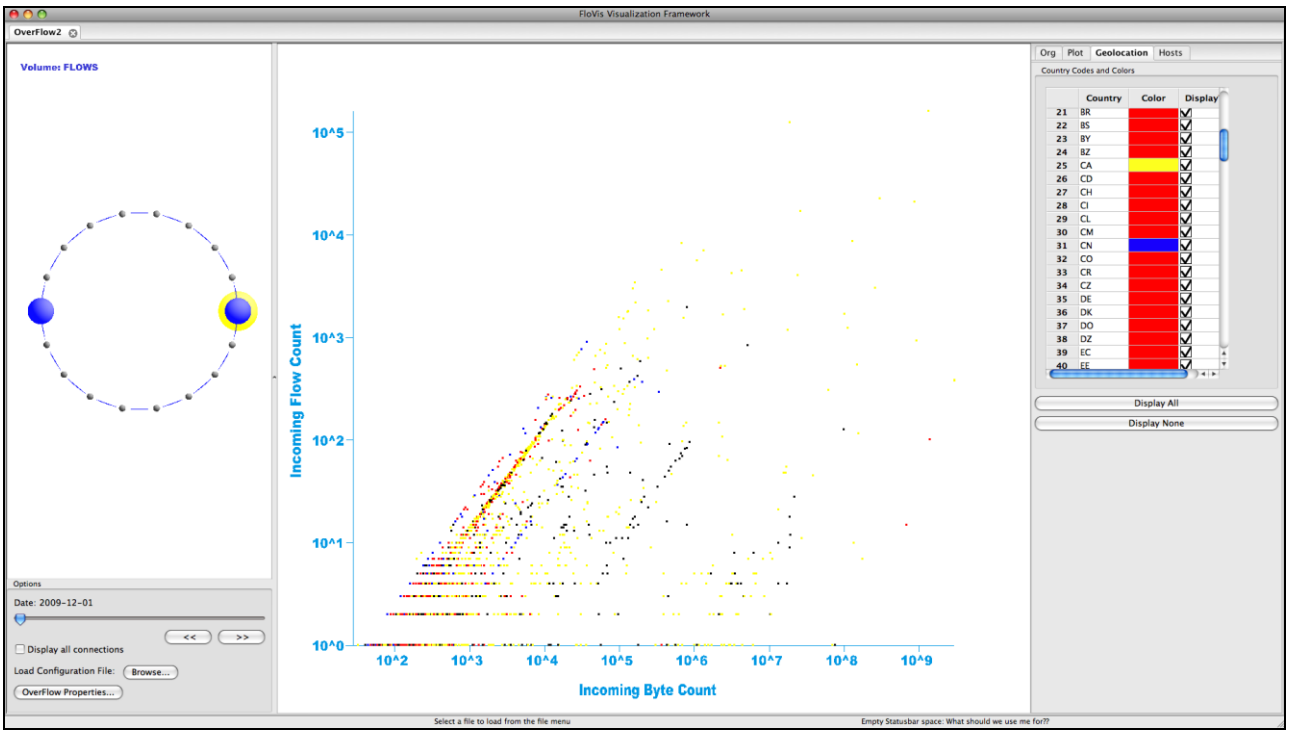

Figure 11: Behavioral feature space; flows vs bytes

The plug-in allows exploration of the host-behaviors by allowing the user to:

- select which attributes should be assigned to the X and Y axes,
- select host colors based on geolocation,
- filter the visible hosts by geolocation and/or individual IP address,
- select a region with the mouse, thus focusing on a set of hosts of interest, and,
- select a set of hosts and immediately drill into the raw flow data as shown in Figure 12.

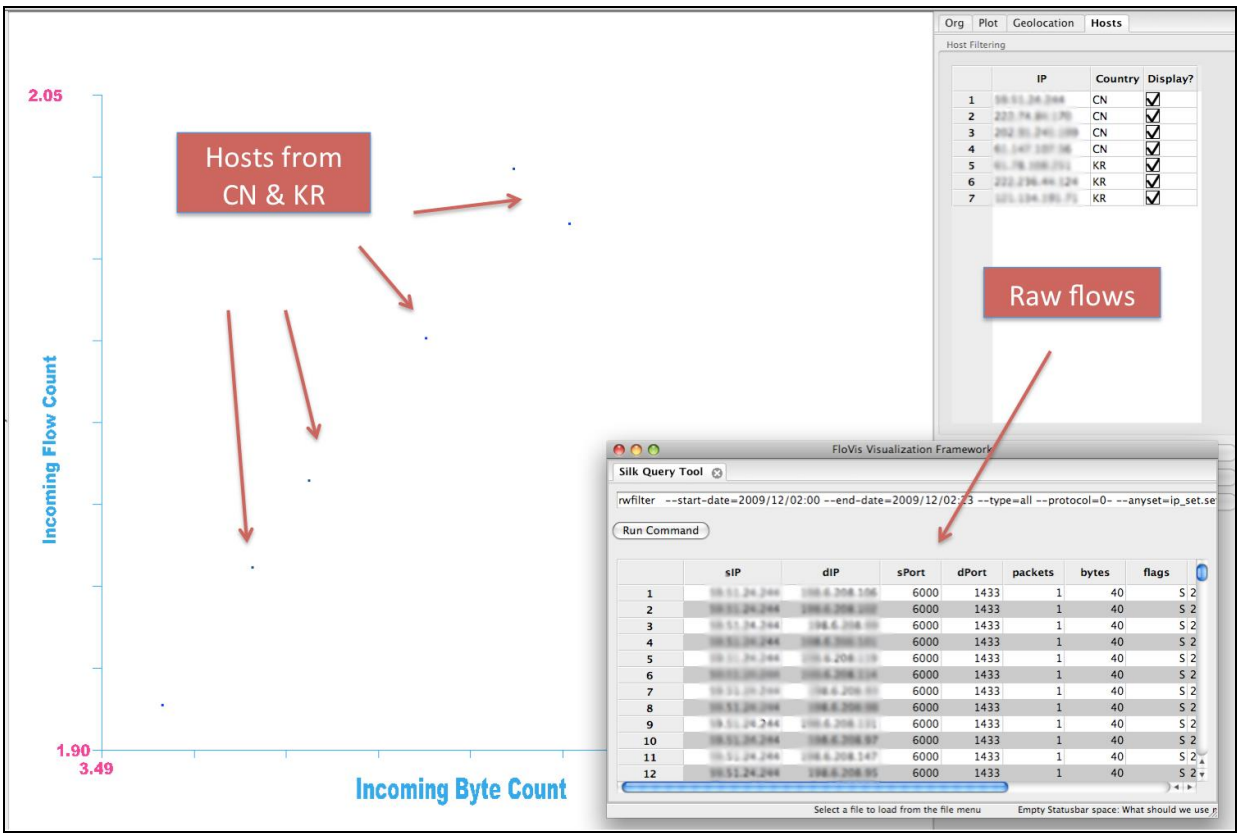

Figure 12: Behavioral feature space; drill down to raw data

3.0 CONCLUSIONS

We have described how the various components of the FloVis framework work in tandem to allow an analyst to visually drill into network data and explore anomalous network behavior. FloVis is an integrated visualization platform designed for rapid visualization development and integration. Visualizations are built as plug-ins allowing the analyst to customize their own visual analysis environment.

Each new plug-in to the framework provides another way of prompting the analyst to ask questions about the data. This can allow one to think laterally with respect to asking questions about network data. Furthermore, new attack paradigms may present themselves as visual anomalies. Since FloVis provides the ability to drill into the raw flow data, we allow the user to more easily determine whether such visual anomalies are worthy of further investigation.

This material is based upon work supported by the Department of Homeland Security under Contract No. N66001-08-C-2032. We also wish to acknowledge the support of Ron McLeod of TARA, CA Labs, and NSERC in this research initiative. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Homeland Security.

4.0 REFERENCES

[1] J. Glanfield, S. Brooks, T. Taylor, D. Paterson, C Smith, C. Gates, J. McHugh. OverFlow: An Overview Visualization for Network Analysis. Accepted to the 6th International Workshop on Visualization for Cyber Security. Atlantic City, NJ. October 11, 2009.

[2] D. Holten. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. IEEE Transactions on Visualization and Computer Graphics, 12(5):741--748, 2006.

[3] B. Johnson and B. Shneiderman. Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In VIS '91: Proceedings of the 2nd conference on Visualization, pp. 284--291, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.

[4] T. Taylor, S. Brooks and J. McHugh. NetBytes Viewer: An Entity-based NetFlow Visualization Utility for Identifying Intrusive Behavior. In Goodall et al. (eds.), Mathematics and Visualization (Proceedings of VizSEC), Springer-Verlag, August, 2008.

[5] T. Taylor, D. Paterson, J. Glanfield, C. Gates, S. Brooks, J. McHugh. FloVis: Flow Visualization System. In Proceedings of the Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH). Washington, DC. March 3-4, 2009.

[6] J. J. van Wijk and E. R. van Selow, E. R. Cluster and Calendar Based Visualization of Time Series Data. In InfoVis '99: Proceedings of the 1999 IEEE Symposium on Information Visualization, pp. 4-9, San Francisco, CA, USA, 1999. IEEE Computer Society Press.

[7] Aggelos Kiayias, Justin Neumann, David Walluck, Owen McCusker, A Combined Fusion and Data Mining Framework for the Detection of Botnets, In Proceedings of the Cybersecurity Applications and Technologies Conference for Homeland Security (CATCH). Washington, DC. March 3-4, 2009.